

Complete amino acid sequence of BSP-A₃ from bovine seminal plasma

Homology to PDC-109 and to the collagen-binding domain of fibronectin

N. G. SEIDAH,* P. MANJUNATH,† J. ROCHEMONT,* M. R. SAIRAM‡ and M. CHRÉTIEN†§

*Biochemical and †Molecular Neuroendocrinology Laboratories and ‡Reproduction Research Laboratory, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, Quebec H2W 1R7, Canada

Bovine seminal plasma was shown to contain three similar proteins, called BSP-A₁, BSP-A₂ and BSP-A₃. Both BSP-A₁ and BSP-A₂ were shown to be molecular variants of a recently characterized peptide called PDC-109. They seem to differ only in their degree of glycosylation and otherwise seem to possess an identical amino acid composition. The work in the present paper deals with the complete characterization of the third member of this series, namely BSP-A₃. The complete amino acid sequence revealed that it is composed of 115 amino acids and predicts a *M_r* of 13403. An analysis of the primary structure of BSP-A₃ revealed a high degree of internal homology, with two homologous domains composed of 39 (residues 28–66) and 43 (residues 73–115) amino acids. An exhaustive computer-bank search for the similarity of this sequence to any known protein, or segment thereof, revealed two significant homologies. The first is between PDC-109 and BSP-A₃, which is so high that we can confidently predict that both proteins evolved from a single ancestral gene. The collagen-binding domain of bovine fibronectin (type II sequence) was also found to be highly homologous to both BSP-A₃ and PDC-109.

INTRODUCTION

Manjunath (1984) isolated and purified (Manjunath & Sairam, 1987) three acidic proteins from bull seminal plasma, designated BSP-A₁, BSP-A₂ and BSP-A₃. These three proteins exhibit both stimulatory and inhibitory actions on the release of pituitary gonadotropins (Manjunath, 1984). The BSP-A₁ and BSP-A₂ proteins turned out to be molecular variants of a previously reported bovine seminal-plasma peptide known as PDC-109 (Esch *et al.*, 1983). These two proteins differ in their degree of glycosylation, but otherwise contain identical polypeptide backbones (Manjunath & Sairam, 1987). Accordingly, BSP-A₁ and BSP-A₂ will be referred to as PDC-109 throughout the text. The third member of this series of analogous proteins, BSP-A₃, has now been fully characterized in terms of its primary sequence and its molecular homology to other known proteins. The work in the present paper allowed us to predict a common ancestral genetic origin for BSP-A₃ and PDC-109. The high degree of homology of both proteins to the collagen-binding domain of fibronectin (type II repeat) also suggests that all three protein domains most probably evolved from a common ancestral gene.

MATERIALS AND METHODS

H.p.l.c. was performed on a Varian 5060 instrument equipped with a Vista 402 plotting integrator and a Beckman 165 UV/Vis detector. The column used was a 5 μ Vydac C₁₈ (0.46 cm \times 25 cm) (The Separations Group, Hesperia, CA, U.S.A.). In all cases the elution system consisted of a linear gradient of acetonitrile in the presence of trifluoroacetic acid (Pierce). Solvent A

consisted of 0.1% (v/v) trifluoroacetic acid in water and solvent B consisted of 0.1% trifluoroacetic acid in acetonitrile (Anachemia). All elutions were performed at a flow rate of 1 ml/min at room temperature, with the linear gradients specified in the legend of the corresponding Figure.

The amino acid compositions of the reduced and carboxymethylated BSP-A₃ (purified as shown in our previous paper; Manjunath & Sairam, 1987), and its CNBr and Endoproteinase Lys-C (Boehringer Mannheim) reaction products, were determined, after either 24 h or 48 h hydrolysis in 5.7 M-HCl *in vacuo* at 110 °C, on a Beckman 121 MB analyser equipped with a Varian Vista 601 plotter/integrator.

The amino acid sequence analysis of the reduced and carboxymethylated BSP-A₃ (1 mg) was performed on a Beckman 890M automatic sequencer, with 3 mg of Polybrene (Aldrich) as a carrier and a 0.33 M-Quadrol program (Seidah *et al.*, 1981), with double coupling at the first cycle and double coupling/cleavage at residues 12, 18, 26, 27, 28 and 32. Amino acid phenylthiohydantoin derivatives were identified by reverse-phase h.p.l.c. as previously described (Seidah *et al.*, 1981; Lazure *et al.*, 1983). The amino acid sequence analysis of the CNBr fragment (residues 97–115 of BSP-A₃) was also performed on the liquid-phase 890M Beckman sequencer. The amino acid sequence analyses of the Endoproteinase Lys-C digestion products were performed on an automated gas-phase sequencer (Applied Biosystems, model 470A). The glass-fibre filter was loaded with 30 μ l of a Biobrene (Applied Biosystems) solution (3 mg) and precycled for four cycles according to the manufacturer's protocol. The h.p.l.c.-purified fraction was then loaded as multiple additions of 30 μ l portions. After loading was

§ To whom correspondence should be addressed.

complete, the sequence was started and the resulting amino acid phenylthiohydantoin were analysed by h.p.l.c., as above.

CNBr cleavage of the reduced and carboxymethylated BSP-A₃ (1 mg) was performed in 70% (v/v) formic acid in the dark for 18 h with a 100:1 molar ratio of CNBr:BSP-A₃ in a total volume of 500 μ l. At the end of the reaction period, the mixture was diluted 10-fold with water and freeze-dried. The resulting powder was then diluted in 6 M-guanidinium chloride/6 M-acetic acid and injected on to a Vydac C₁₈ column. Elution was performed with an acetonitrile gradient in 0.1% trifluoroacetic acid as described above.

Digestion of the CNBr fragment 1–96 of BSP-A₃ with Endoproteinase Lys-C was performed for 4.5 h at 37 °C, with a 1:20 (w/w) enzyme/substrate ratio. The products were then immediately injected on to the h.p.l.c. column and purified. Amino acid analysis of the various fragments so obtained allowed the identification of these peptides. The sequence of some of them was performed in order to confirm the proposed primary structure.

In order to screen for sequence homology of BSP-A₃ with any known protein or segment thereof, we used the protein data bank available on-line from the Protein Identification Resource (National Biomedical Foundation, Georgetown University, Washington, DC, U.S.A.). The programs MATCH, ALIGN, RELATE and PRPLOT, using the Kyte & Doolittle (1982) 'Hydrophathy Index' or the Rose *et al.* (1985) 'Fractional Exposure Index', were all used.

RESULTS

Amino acid sequence determination of BSP-A₃

Preliminary runs were first initiated in order to probe the ease with which the reduced and carboxymethylated BSP-A₃ was amenable to sequence analysis. For this we used 7 nmol of peptide and sequenced 59 residues. This preliminary run permitted us to locate the position of the first three proline residues and potential problems during the sequencing program used. This is why on the second sequencing trial we were able to extend the *N*-terminal sequence to 100 residues (Fig. 1a). The strategy involved the use of double coupling and double cleavage at residues 12, 18, 26, 27, 28 and 32. The proline residues at positions 12, 18 and 32 were causing excessive carry-over, owing to slow cleavage of the thiocarbamoyl derivative, and double coupling/cleavage lessened this problem. The same procedure was also used to decrease the excessive carry-over caused by the stretch Lys-Asp-Asn-Lys, occupying positions 25–28, which was also found to be sequenced with difficulty, with a fall in yield after Asp-26. A similar phenomenon was observed during the sequence of β -Inhibin in the stretch Asp-Asn-Cys (Seidah *et al.*, 1984b).

This sequence determination of the first 100 residues of BSP-A₃ greatly facilitated the completion of the structural characterization of this molecule. Since BSP-A₃ contains only one methionine (Table 1), at residue no. 96 (Fig. 1a), the use of CNBr generated two fragments (Fig. 2a). The fragment residues 1–96 (Table 1) was used for further digestion with Endoproteinase Lys-C (Fig. 2b), and the other fragment was sequenced to completion (Fig. 1b). The amino acid analysis of this

C-terminal CNBr fragment (Table 1) and the sequence data agree on the length of this peptide as being 19 amino acids (residues 97–115). Since the sequence Met-Asn-Tyr-Trp-Cys- was established from the extended 100 residues run (Fig. 1a), this provided a four-amino-acids sequence overlap with the C-terminal CNBr fragment 97–115 (which starts with the sequence Asn-Tyr-Trp-Cys-). However, the C-terminal Cys-115 was found to be resistant to the action of carboxypeptidase Y (Manjunath & Sairam, 1987). Interestingly, both BSP-A₃ fragments 1–96 and 97–115 when separated by reverse-phase h.p.l.c. are eluted in multiple peaks, all possessing similar amino acid compositions (Fig. 2a). Since BSP-A₃ is not glycosylated (Manjunath & Sairam, 1987), we believe that this heterogeneity is partly due to various oxidation states of tryptophan residues, which are not quantified in our amino acid analysis. Possibly oxidation of some cysteine residues could also contribute to this heterogeneity.

Since the extended 100-amino-acid sequence of BSP-A₃ did not permit the unambiguous identification of residues 81 and 86 (Fig. 1a), we chose to cleave the residues-1–96 CNBr fragment further with Endoproteinase Lys-C proteinase, which cleaves C-terminal to lysine residues. The products of this reaction were purified by h.p.l.c. as shown in Fig. 2(b). Many peptides were generated, and the identity of some of them was confirmed by amino acid analysis and sequence. This method allowed the generation of two key peptides, namely fragments residues 74–83 and 84–90. The sequences of these peptides allowed the affirmation of the missing residues 81 and 86 as glutamic acid and aspartic acid respectively.

Accordingly, the proposed primary structure of bovine seminal plasma BSP-A₃ is presented in Fig. 3, together with the strategy used for its complete sequence determination. This 115-amino-acid BSP-A₃ sequence predicts an *M_r* of 13403. It is clear that BSP-A₃, like PDC-109 (Esch *et al.*, 1983), is composed of two structurally similar domains, A (residues 28–66) and B (residues 73–115), of 39 and 43 amino acids respectively. It is quite possible, by analogy to PDC-109 (Esch *et al.*, 1983), that each domain contains two disulphide bridges.

Sequence homology to PDC-109 and fibronectin

In order to assess the likelihood of sequence similarity of BSP-A₃ to any known protein or segment thereof, an exhaustive computer data-bank search using the program MATCH revealed only two proteins with significant similarities, namely PDC-109 (Esch *et al.*, 1983) and the collagen-binding domain (type II sequence) of bovine plasma fibronectin (Petersen *et al.*, 1983). Next, by using the program ALIGN, which utilizes a mutation data matrix, the best alignment between these three molecules is shown in Fig. 4. It is clear that a very high degree of sequence similarity exists between these three proteins, especially between BSP-A₃ and PDC-109. In fact, the best alignment score was obtained for the pair BSP-A₃/PDC-109, with 27.99 s.d. units from random with 72 identities out of a possible 109 matches between residues and three breaks (see Table 2). This translates into a 5×10^{-195} probability of getting this score by chance. The alignment of the pairs BSP-A₃/fibronectin and PDC-109/fibronectin gave a 5×10^{-31} and 3×10^{-28} probability that these homologies would occur by chance. The scores obtained by using the program

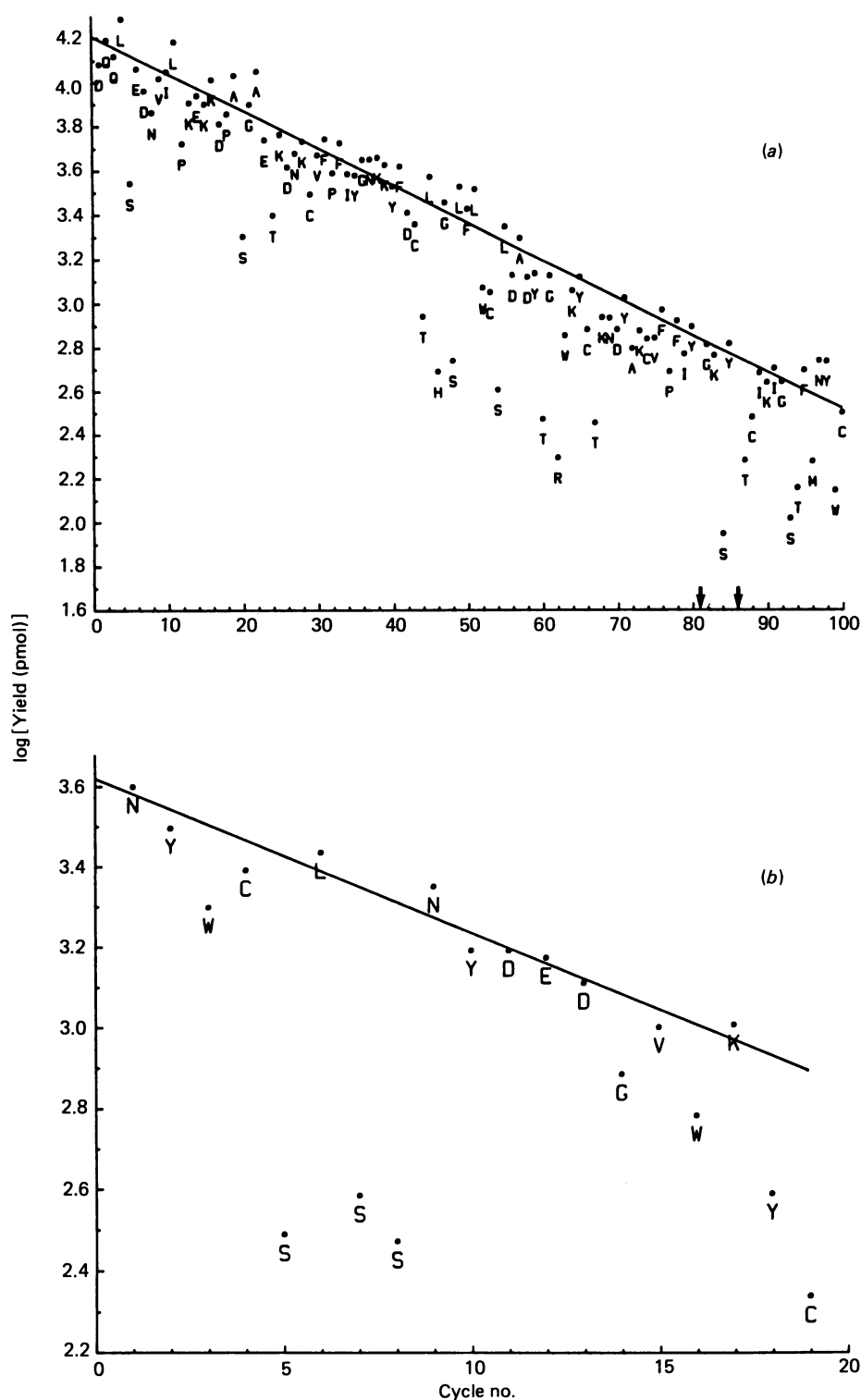


Fig. 1. Plot of the logarithm of the yield of each amino acid phenylthiohydantoin versus the cycle number obtained during the automatic sequencing of (a) the reduced and *S*-carboxymethylated BSP-A₃ and (b) the CNBr fragment residues 97–115 (see Table 1 and Fig. 2a)

In (a) the arrows indicate the positions of the two unidentified residues 81 and 86, which were later affirmed to be Glu and Asp respectively. Residue 81 could not be unambiguously identified as Glu in view of the high Glu background observed at this stage at this extended sequence. Residue 86 was lost owing to mechanical failure during the automatic phenylthiohydantoin conversion. The initial yield, repetitive yield and correlation coefficient obtained by linear-regression analysis were respectively as follows: (a) 16207 pmol, 96.2%, 0.9671; (b) 3936 pmol, 91.6%, 0.9576. Remarkably, with an initial yield of 16207 pmol of BSP-A₃ and a repetitive yield of 96.1% we were able to identify the 98 *N*-terminal residues out of 100 (a). Furthermore, the *C*-terminal CNBr fragment representing the segment 97–115 of BSP-A₃ was sequenced to completion with 3936 pmol of peptide with a repetitive yield of 91.6% (b).

Table 1. Amino acid analysis of reduced and carboxymethylated BSP-A₃ and its CNBr fragments (see Fig. 2)

Here we also compare the values reported for BSP I (Esch *et al.*, 1983), a peptide we believe to be identical with BSP-A₃. The calculated amino acid composition based on our sequence data is also shown. These data predict an M_r of 13403 for the 115-amino-acid BSP-A₃. The values in parentheses for the CNBr fragments represent the expected numbers based on the proposed sequence. Abbreviation: N.D., not determined.

| Amino acid | BSP-A ₃ † (the present work) | BSP I (Esch <i>et al.</i> , 1983) | BSP-A ₃ sequence | CNBr peptides‡ | |
|------------|--|--------------------------------------|--------------------------------|----------------|-----------------|
| | | | | Residues 1–96 | Residues 97–115 |
| Asx | 17.0 | 13.7 | 17 | 13.0 (13) | 4.0 (4) |
| Thr | 5.8 | 5.2 | 6 | 5.7 (6) | 0.0 (0) |
| Ser | 8.5 | 8.3 | 9 | 4.8 (6) | 2.4 (3) |
| Glx | 7.7 | 7.4 | 7 | 6.3 (6) | 1.5 (1) |
| Pro | 3.6 | 4.9 | 4 | 5.0 (4) | 0.0 (0) |
| Gly | 7.7 | 6.8 | 7 | 5.7 (6) | 1.5 (1) |
| Ala | 3.9 | 4.1 | 4 | 3.9 (4) | 0.0 (0) |
| Cys* | 7.9 | 6.9 | 8 | 5.2 (6) | 1.9 (2) |
| Val | 4.1 | 3.3 | 4 | 3.0 (3) | 1.0 (1) |
| Met | 1.0 | 1.0 | 1 | N.D. (1)§ | 0.0 (0) |
| Ile | 4.9 | 4.4 | 5 | 5.4 (5) | 0.0 (0) |
| Leu | 7.4 | 7.1 | 7 | 6.3 (6) | 1.2 (1) |
| Tyr | 10.0 | 10.1 | 10 | 7.9 (7) | 2.9 (3) |
| Phe | 7.8 | 6.9 | 7 | 6.8 (7) | 0.0 (0) |
| His | 1.0 | 1.0 | 1 | 1.1 (1) | 0.0 (0) |
| Lys | 12.6 | 12.3 | 13 | 11.0 (12) | 1.4 (1) |
| Arg | 1.4 | 1.2 | 1 | 1.3 (1) | 0.0 (0) |
| Trp | 5.0 | 4.3 | 4 | N.D. (2) | N.D. (2) |

* Determined as *S*-carboxymethylcysteine.

†,‡ Amino acid analysis after 24 h or 48 h hydrolysis respectively.

§ Present as homoserine (lactone).

RELATE (Table 3) also confirm the high degree of homology between these proteins. All these statistical calculations strongly suggest that these three sequences arose from a single ancestral gene.

Hydropathy-index distribution

In view of the high degree of sequence homology found between each individual domain A and B and between the corresponding domain in BSP-A₃, PDC-109 or the collagen-binding domain of fibronectin, it was decided to evaluate the hydropathy index (Kyte & Doolittle, 1982) and the residue-exposure index (Rose *et al.*, 1985) distribution of each domain individually, in order to possibly identify common structural 'motifs' (Seidah *et al.*, 1986). The side-chain hydropathy (Kyte & Doolittle, 1982), a method of detecting common tertiary-structural patterns of solvent-exposed and buried residues in related proteins, revealed the best similarities (Fig. 5). Essentially the same overall pattern was obtained by using the residue exposure index (Rose *et al.*, 1985), a measure of average solvent exposure determined from protein crystallographic data (results not shown). Fig. 5 shows that, as expected from the ALIGN (Table 2) and RELATE (Table 3) data, BSP-A₃ and PDC-109 are the most hydropathically homologous pair in either domains A or B.

DISCUSSION

Within the last few years, the amino acid sequences of a number of novel polypeptides exhibiting Inhibin-like

activity have been determined (Esch *et al.*, 1983; Seidah *et al.*, 1984a,b; Li *et al.*, 1985; Mason *et al.*, 1985, 1986; Forage *et al.*, 1986). The cDNA structures coding for some of these peptides have now been elucidated (Mason *et al.*, 1985, 1986; Forage *et al.*, 1986). From this work it became apparent that the protein sequences obtained for peptides derived from either seminal plasma, such as human α -Inhibin (Seidah *et al.*, 1984a; Li *et al.*, 1985; Lilja & Jeppsson, 1985), human β -Inhibin (Seidah *et al.*, 1984b), bovine PDC-109 (Esch *et al.*, 1983), or rat seminal-vesicle secretory protein SVS-IV (Yu-Ching *et al.*, 1980; Mansson *et al.*, 1981), are not related to the follicular-fluid Inhibin (Mason *et al.*, 1985, 1986; Forage *et al.*, 1986). Furthermore, these male-derived peptides are not derived from the testis Sertoli cells. Rather, α -Inhibin is found mostly in seminal vesicles (Lilja & Jeppsson, 1985), and β -Inhibin is mostly found in prostate epithelial cells (Beksac *et al.*, 1984). The exact physiological role of these polypeptides is not yet fully understood.

The data in the present work and in the preceding paper (Manjunath & Sairam, 1987) clearly demonstrate that bovine seminal plasma contains two proteins, PDC-109 (Esch *et al.*, 1983) and BSP-A₃, exhibiting a very high degree of sequence homology, suggesting that they arose via a gene-duplication event from a common ancestral gene. Whether BSP-A₃ and PDC-109 are found within a common precursor molecule or originate from two different mRNAs has not yet been determined. However, the tissue from which both peptides originate has now clearly been defined as the seminal vesicles (P. Manjunath, unpublished work). So far, four different

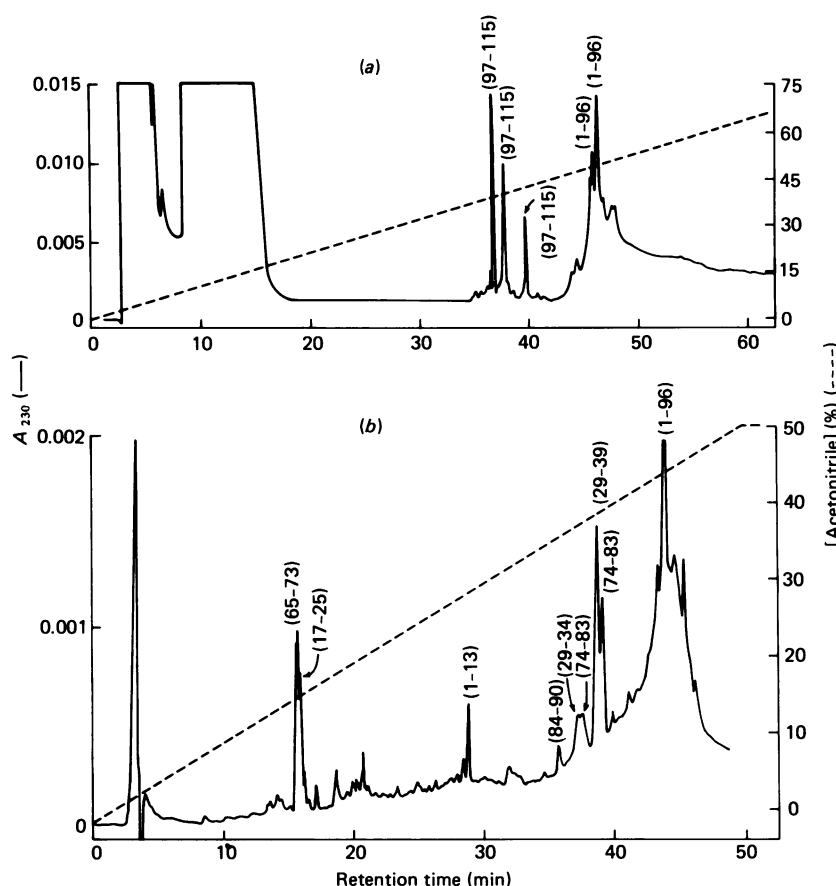


Fig. 2. Reverse-phase h.p.l.c. purification of (a) the CNBr digest of *S*-carboxymethylated BSP-A₃ and (b) the Endoproteinase Lys-C digest of the CNBr fragment 1-96 of BSP-A₃ obtained from (a)

The identity of the peptides shown above some peaks is based on their amino acid composition (see Table 1) and their sequence. In (b) the unidentified peaks probably contain peptides spanning the segment residues 40-64. The heterogeneity in the CNBr fragments 1-96 and 97-115 (a) and the Endoproteinase Lys-C fragments 29-39 and 74-83 (b) is probably due to different oxidation states of the Trp and/or Cys residues. The linear acetonitrile/trifluoroacetic acid gradient used is shown with each chromatogram as a broken line. The column used was a 5 μ Vydac C₁₈ column (0.46 cm \times 25 cm).

protein sequences were identified as originating from seminal vesicles; these include human α -Inhibin (Seidah *et al.*, 1984a; Li *et al.*, 1985; Lilja & Jeppsson, 1985), rat seminal-vesicle secretory protein SVS-IV (Yu-Ching *et al.*, 1980; Mansson *et al.*, 1981), PDC-109 and BSP-A₃ (P. Manjunath, unpublished work).

The movement of cells to reach their destination unerringly is a highly organized event. The best-understood mechanism by which this organization of cells is maintained is the one involving versatile anchoring and organizing proteins known as fibronectins, and collectively as fibronectin. This molecule is a dimer consisting of two similar subunits held together by disulphide bridges. In human fibronectin each subunit is composed of about 2355 amino acids (Kornblihtt *et al.*, 1985). However, differential RNA splicing may give rise to at least 12 versions of each subunit of this protein, varying in length between 2145 and 2445 amino acids (Hynes, 1986). The action of proteolytic enzymes revealed that each subunit of fibronectin can be subdivided into domains, within which the protein chain is tightly folded, and hence resistant to degradation. Individual domains account for fibronectin's ability to bind to various proteins and surface membranes. The

amino acid sequence of fibronectin reveals unique structures within the domains. These contain small repeated protein modules, whose similarities of sequence allow them to be classified into three types (I, II and III). Type I sequences are found in the domain responsible for the ability of fibronectin to bind fibrin. Type II is only found in the domain which binds collagen, and type III is present in the cell-, the heparin- or the fibrin-binding domains (Hynes, 1986).

Of interest to us in the present paper is the collagen-binding domain containing type II sequences. It is those sequences which are homologous to BSP-A₃ and PDC-109. Baker (1985) was the first to notice the high degree of homology that exists between PDC-109 and the type-II sequence of fibronectin. However, our statistical data, although in general agreement with his reported significant homology between PDC-109 and type II sequences, did not detect any significant correlation between either PDC-109 or BSP-A₃ with the segment 225-275 of tissue plasminogen activator t-PA (Banyai *et al.*, 1983). In fact, the ALIGN score obtained with this segment was -0.36 s.d. units. We believe the discrepancy to be due to his use of the s.d. in his calculation of the probability of relatedness rather than the more suitable

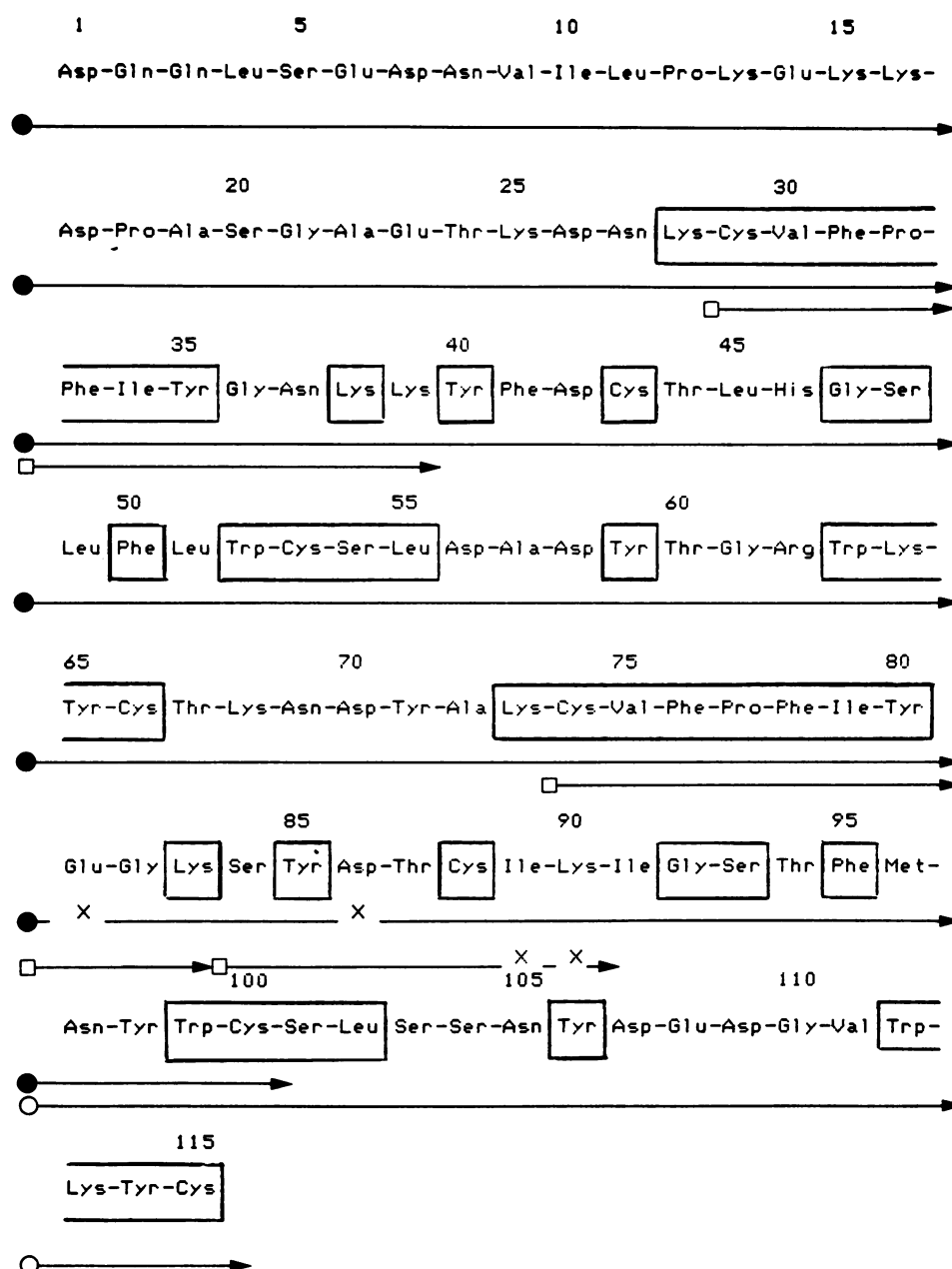


Fig. 3. Proposed primary structure of bovine seminal-plasma 115-amino-acid BSP-A₃.

The complete sequence was obtained via the sequence analysis of intact *S*-carboxymethylated BSP-A₃ (●), CNBr fragment 97–115 (○) and Endoproteinase Lys-C fragments (□): X indicates a residue not identified during the sequence analysis. Boxed amino acids denote homologous sequences between domains A (residues 28–66) and B (residues 73–115).

Table 2. Alignment scores obtained with the program ALIGN for the peptides BSP-A₃ (the present work), PDC-109 (Esch *et al.*, 1983) and the type II sequence of bovine fibronectin (collagen-binding domain) (Petersen *et al.*, 1983)

The scores are expressed in z-units of standard deviation (S.D.) from the mean (Dayhoff *et al.*, 1983). The program uses a mutation data matrix, and the best alignment obtained should give the highest z-value with the minimum number of gaps (breaks) and maximum number of identical matches between a pair of sequences. Also shown in this table is the calculated probability (*P*) that such a homology is due to chance, based on the z-value obtained. Obviously, BSP-A₃ and PDC-109 share the highest alignment score, with an infinitesimal chance that such a homology is due to chance ($P < 5 \times 10^{-195}$).

| Protein pair aligned | z-units S.D. | Identical matches | Breaks | Chance probability |
|-----------------------------|--------------|-------------------|--------|------------------------|
| BSP-A ₃ /PDC-109 | 27.99 | 72/109 | 3 | $< 5 \times 10^{-195}$ |
| BSP-A ₃ /type II | 11.5 | 37/111 | 7 | $< 5 \times 10^{-31}$ |
| PDC-109/type II | 10.95 | 37/107 | 6 | $< 3 \times 10^{-28}$ |

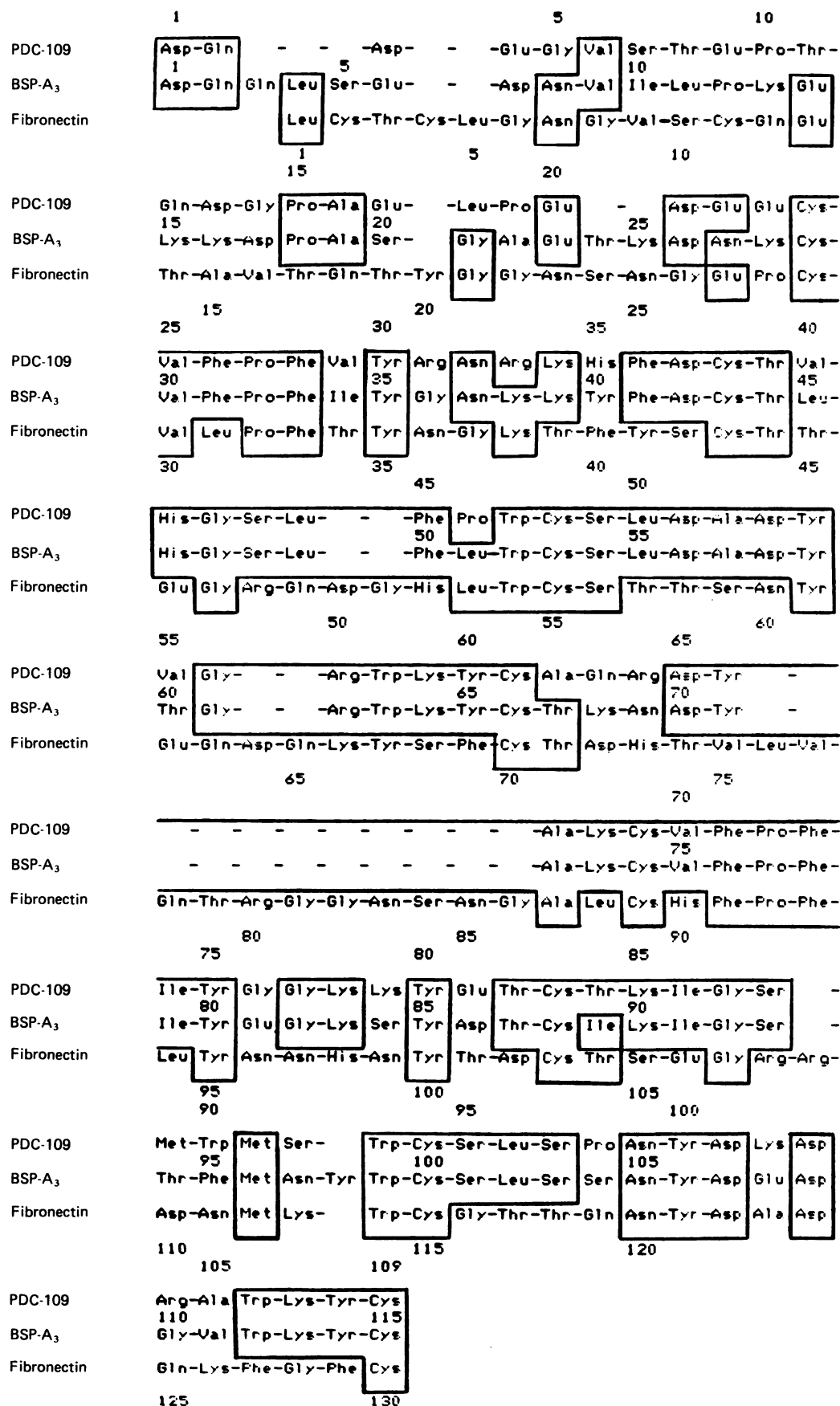


Fig. 4. Optimal alignment of BSP-A₃, PDC-109 (Esch *et al.*, 1983) and the collagen-binding domain of bovine fibronectin (Petersen *et al.*, 1983) based on the program ALIGN from the National Biomedical Research Foundation (Washington, DC)

With this program and a mutation data matrix for the calculation of the best scores, gaps were introduced to maximize the alignment. The alignment scores are given in Table 2. Identically positioned residues are boxed.

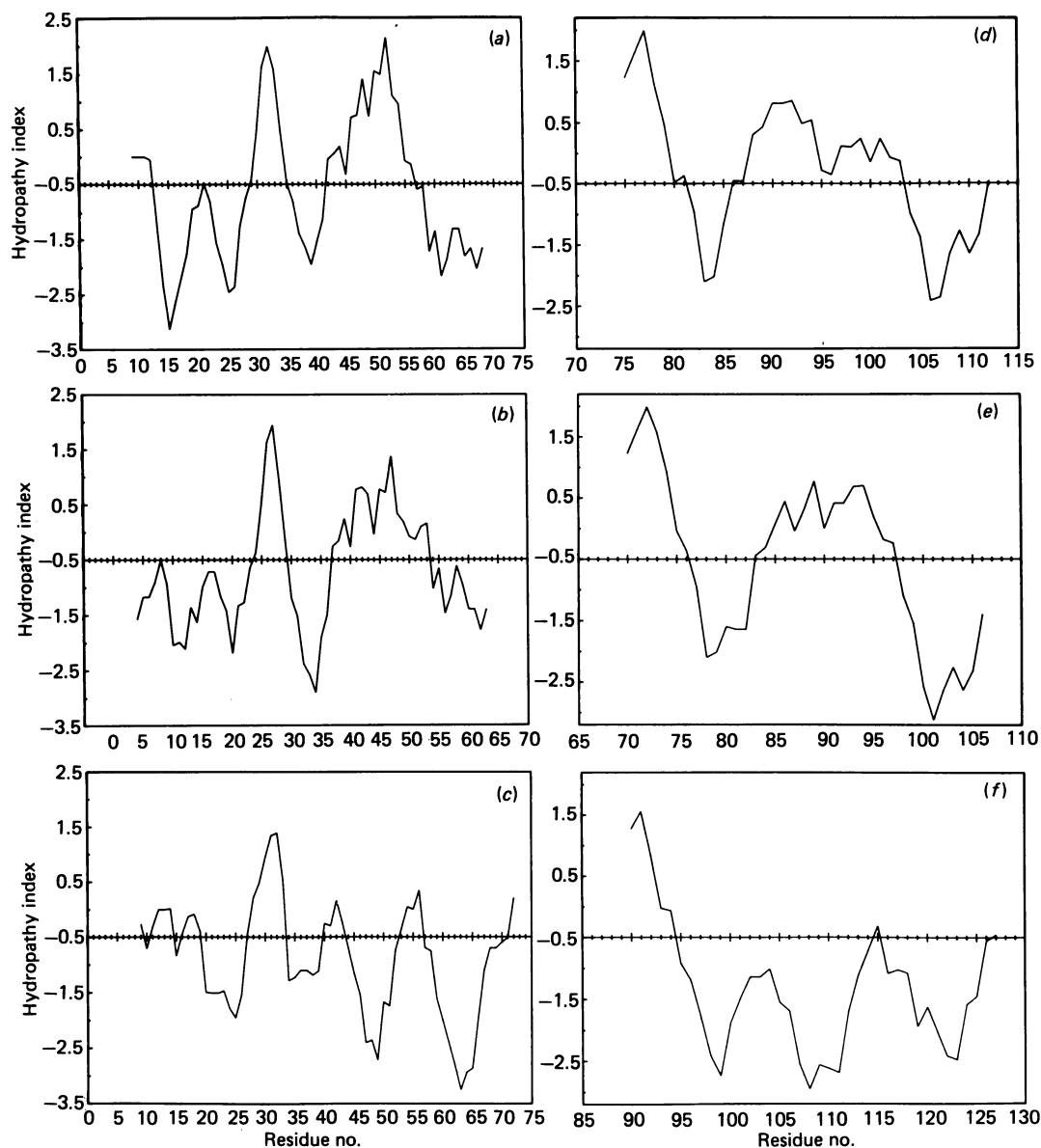


Fig. 5. Hydropathy-index distribution

By using the 'hydropathy index' for each amino acid (Kyte & Doolittle, 1982), hydropathy values were calculated for each residue taken as a running average with the preceding and following three residues (window = 7 amino acids span). These hydropathy values are then plotted versus the residue number. Maxima and minima correspond to predicted buried and exposed regions respectively. Accordingly, we have compared the hydropathy index distribution for each of the domains A and B separately. On the left is shown the hydropathy-index distribution of domain A of BSP-A₃ (a, residues 7–71), PDC-109 (b, residues 1–66) and the collagen-binding domain of fibronectin (c, residues 6–75). On the right is depicted the hydropathy-index distribution of domain B of BSP-A₃ (d, residues 72–115), PDC-109 (e, residues 67–109) and the collagen-binding domain of fibronectin (f, residues 87–130). Clearly a high degree of hydropathic homology exists between the corresponding domains of these three molecules. Similar conclusions were drawn when we used the 'residue exposure' (Rose *et al.*, 1985) instead of the 'hydropathy index' distribution (results not shown).

z-factor, defined as the difference of the total score from the mean score in S.D. units (Dayhoff *et al.*, 1983). Furthermore, t-PA is more related to type I rather than type II repeats in fibronectin (Banyai *et al.*, 1983).

It has been reported that the collagen-binding domain of fibronectin promotes viral transformation of chicken fibroblasts in culture by Rous sarcoma virus (DePetro *et al.*, 1981). Interestingly, Baker (1985) suggested that PDC-109 could be one of the predisposing factors for the subsequent development of Acquired Immune Deficiency

Syndrome (AIDS). We believe that it is still too early to define the exact role of these proteins *in vivo*. However, their strong degree of homology to the collagen-binding domain of fibronectin suggests a homologous role as a biological organizer of cells in the seminal plasma, possibly guiding their migration.

Remarkably, BSP-A₃ and PDC-109 were isolated on the basis of their ability to release gonadotropins by rat anterior-pituitary cell cultures (Manjunath, 1984; Manjunath & Sairam, 1987) or by their ability to antag-

Table 3. Scores and probability of chance relationship between peptides BSP-A₃, PDC-109 and the type II sequence of fibronectin calculated with the program RELATE

The definitions of the terms are as given in Table 2. Again with this statistical program the probability that BSP-A₃, PDC-109 and the collagen-binding domain sequences are only related by chance is very low indeed, with BSP-A₃ and PDC-109 being the closest relatives.

| Protein pair related | z-units S.D. | Chance probability |
|-----------------------------|-----------------|------------------------|
| BSP-A ₃ /PDC-109 | 36.05 | $< 7 \times 10^{-285}$ |
| BSP-A ₃ /type II | 17.76 | $< 7 \times 10^{-71}$ |
| PDC-109/type II | 8.84 | $< 5 \times 10^{-19}$ |

onize the effect of luteinizing-hormone-releasing hormone (Esch *et al.*, 1983). These effects *in vitro*, however, are observed at relatively high concentrations (in the micromolar range), which would rather suggest that these activities are pharmacologically rather than physiologically significant. Obviously much more work needs to be done in order to unravel the biological function(s) of this homologous series of proteins found in seminal plasma.

We thank Mrs. Josée Hamelin and Mr. Gilles De Serres for their expert technical assistance and Mrs. Diane Laliberté for secretarial expertise. We also thank the J.A. De Séve Foundation, the Medical Research Council of Canada, Fonds de la Recherche en Santé du Québec and the Mellon Foundation for their financial support. P.M. is the recipient of a scholarship from the Fonds de la Recherche en Santé du Québec.

REFERENCES

- Baker, M. E. (1985) *Biochem. Biophys. Res. Commun.* **130**, 1010–1014
- Banyai, L., Varadi, A. & Patthy, L. (1983) *FEBS Lett.* **163**, 37–41
- Beksac, M. S., Khan, S. A., Eliasson, R., Skakkebaek, N. E., Sheth, A. R. & Diczfalusy, E. (1984) *Int. J. Androl.* **21**, 695–700
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545
- DePetro, G., Barlati, S., Vartio, T. & Vaher, A. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4965–4969
- Esch, F. S., Ling, N. C., Bohlen, P., Ying, S. Y. & Guillemin, R. (1983) *Biochem. Biophys. Res. Commun.* **113**, 861–867
- Forage, R. G., Ring, J. M., Brown, R. W., McInerney, B. V., Cobon, G. S., Gregson, R. P., Robertson, D. M., Morgan, F. J., Hearn, M. T. W., Findlay, J. K., Wettenthal, R. E. H., Burger, H. G. & De Kretser, D. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3091–3095
- Hynes, R. O. (1986) *Sci. Am.* **254**, 42–51
- Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4**, 1755–1759
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Lazure, C., Seidah, N. G., Chrétien, M., Lallier, R. & St. Pierre, S. (1983) *Can. J. Biochem. Cell Biol.* **61**, 287–292
- Li, C. H., Hammonds, R. G., Jr., Ramasharma, K. & Chung, D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4041–4044
- Lilja, H. & Jeppsson, J.-O. (1985) *FEBS Lett.* **182**, 181–184
- Manjunath, P. (1984) in *Gonadal Proteins and Peptides and their Biological Significance* (Sairam, M. R. & Atkinson, L. E., eds.), pp. 49–61, World Scientific Publishing Co., Singapore
- Manjunath, P. & Sairam, M. R. (1987) *Biochem. J.* **241**, 685–692
- Mansson, P.-E., Sugino, A. & Harris, S. E. (1981) *Nucleic Acids Res.* **9**, 935–946
- Mason, A. J., Hayflick, J. S., Ling, N., Esch, F., Ueno, N., Ying, S.-Y., Guillemin, R., Niall, H. & Seeburg, P. H. (1985) *Nature (London)* **318**, 659–663
- Mason, A. J., Niall, H. D. & Seeburg, P. H. (1986) *Biochem. Biophys. Res. Commun.* **136**, 957–964
- Petersen, T. E., Thøgersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 137–141
- Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985) *Science* **229**, 834–838
- Seidah, N. G., Rochemont, J., Hamelin, J., Lis, M. & Chrétien, M. (1981) *J. Biol. Chem.* **256**, 7977–7984
- Seidah, N. G., Ramasharma, K., Sairam, M. R. & Chrétien, M. (1984a) *FEBS Lett.* **167**, 98–102
- Seidah, N. G., Arbatti, N. J., Rochemont, J., Sheth, A. R. & Chrétien, M. (1984b) *FEBS Lett.* **175**, 349–355
- Seidah, N. G., Donohue-Rolfe, A., Lazure, C., Auclair, F., Keusch, G. T. & Chrétien, M. (1986) *J. Biol. Chem.* **261**, 13928–13931
- Yu-Ching, E. P., Silverberg, A. B., Harris, S. E. & Li, S. S.-L. (1980) *Int. J. Peptide Protein Res.* **16**, 143–146

Received 18 July 1986/6 October 1986; accepted 3 December 1986